**Paper No: PU-SOE-CSE-41**

**Category Classification of the Training Set Combined with Sentence Multiplication for Semantic Data Extraction Using GENI Algorithm**

Saravana Kumar Coimbatore Shanmugama, Santhosh Rajendranb, Amudhavalli Padmanabhanc, **Kalaiarasan Chellan****d**.

a,b,c. Department of CSE, Karpagam Academy of Higher Education, Coimbatore, India

d.School of Engineering, Presidency University, Bengaluru, India

**Abstract**

Background: Increase in the internet data has increased the priority in the data extraction accuracy. Accuracy here lies with what data the user has requested for and what has been retrieved. The same large data sets that need to be analyzed make the required information retrieval a challenging task.

Objective: To propose a new algorithm in an improved way than the traditional methods to classify the category or group to which each training sentence belongs.

Method: Identifying the category to which the input sentence belongs is achieved by analyzing the Noun and Verb of each training sentence. NLP is applied to each training sentence and the group or category classification is achieved using the proposed GENI algorithm so that the classifier is trained efficiently to extract the user requested information.

Results: The input sentences are transformed into a data table by applying GENI algorithm for group categorization. Plotting the graph in R tool, the accuracy of the group extracted by the Classifier involving GENI approach is higher than that of Naive Bayes & Decision Trees.

**Keywords:**

Text Classification, Semantic Association, Supervised Learning Text Classification, Semantic Association, Supervised Learning

**Publication Details:**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Journal Name** | **Vol.** | **Month & Year**  | **Page No.** | **Publisher** | **Scimago Ranking** |
| Recent Patents on Computer Science | 13(4) | June. 2020 | 588-594 | Bentham Sciences Publishers B.V | Q4 |